DOCUMENT RESUME

ED 275 762 TM 860 679

AUTHOR Huntley, Renee M.; Carlson, James E.
TITLE The Effect of Item Presentation on Item Difficulty and Discrimination in Language-Usage Tests.
PUB DATE Apr 86
NOTE 22p.; Paper presented at the Annual Meeting of the National Council on Measurement in Education (San Francisco, CA, April 17-20, 1986).
PUB TYPE Speeches/Conference Papers (150) -- Reports - Research/Technical (143)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS College Entrance Examinations; *Difficulty Level; Goodness of Fit; *Item Analysis; *Language Tests; Language Usage; Mathematical Models; Multiple Choice Tests; Secondary Education; *Test Construction; *Test Format; *Test Items; Test Validity
IDENTIFIERS *American College Testing Program; Chi Square

ABSTRACT

This study compared student performance on language-usage test items presented in two different formats: as discrete sentences and as items embedded in passages. American College Testing (ACT) Program's Assessment experimental units were constructed that presented 40 items in the two different formats. Results suggest item presentation may not affect item difficulty to any significant degree, even though some language-usage classifications may be particularly susceptible to the impact of format. Questions about content validity are also considered. Samples of the two formats, passage presentation and discrete-sentence presentation, are included in the appendices. (Author/JAZ)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED275762

The Effect of Item Presentation on Item Difficulty and Discrimination in Language-Usage Tests

Renee M. Huntley
American College Testing Program

James E. Carlson
American College Testing Program

ABSTRACT

This study compared student performance on language-usage test items presented in two different formats: as discrete sentences and as items embedded in passages. ACT Assessment experimental units were constructed that presented forty items in the two different formats. Results suggest item presentation may not affect item difficulty to any significant degree, even though some language-usage classifications may be particularly susceptible to the impact of format. Questions about content validity are also considered.

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☑ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

R. M. Huntley

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

The Effect of Item Presentation on Item Difficulty in Language-Usage Tests

INTRODUCTION

Test constructors seem to be partial to multiple-choice instruments in the assessment of language skills because these tests seem to do "an amazingly effective job of ordering students in the same way as a trained group of readers would after reading a sizable sample of their actual writing" (Godshalk, Swineford & Coffman, 1966, p. 21). Furthermore, multiple-choice tests can provide useful information for given purposes reliably and efficiently in terms of administration and money (Huntley, Schmeiser & Stiggins, 1979, pp. 13-14).

Multiple-choice language tests can be constructed in varying formats. One format embeds items within a passage: a continuous piece of discourse with underlined and numbered portions. Alternative responses are provided, and the examinee must choose the one alternative that, for example, best expresses the idea, makes the statement appropriate for standard written English, or is worded most consistently with the style and tone of the passage as a whole. A second format merely presents discrete sentences, each of which contains a particular language situation that is being tested (e.g., subject-verb agreement, use of the comma for coordinate adjectives, etc.). The differences between these two formats are significant enough to raise many questions for the test constructor and the test user. Nevertheless, although there are many instances in the literature of studies comparing "direct" and "indirect" language assessments (i.e., multiple-choice tests versus writing samples), a recent literature search could uncover no instance of studies comparing these two methods of item presentation.

Test constructors, particularly among English teachers, very often choose to construct a language test in the passage-related format precisely because it seems to address the concerns of their discipline — rhetorical concerns — better than a format of discrete sentences might: without the context of an essay, how could one evaluate appropriateness, patterns, relationships of word choice to meaning, and so forth? After all, one of the fundamental language tenets is that language "situations" not be wrested from the context of "purpose, audience, intention, and occasion" (Huntley, 1979, p. 17). The passage-related format, however, with its heavier reading load, may make greater demands on the examinee than the discrete-sentence format. On the other hand, there may be serious limitations to the kinds of language situations that can be assessed by discrete sentences. These concerns raise further issues regarding test limitations, test efficiency, test purpose, content validity.

METHOD

This study investigated the effect on item difficulty of language-usage test items presented in two different formats: as discrete sentences and as items embedded in passages. Two experimental tryouts, Tests A and B, were constructed that each contained twenty items. These forty items had been culled from the thirteen passage-related tryout units in which they had originally been tried out. The items selected could stand alone coherently within the experimental tryouts, where they were randomly ordered. The content classification of these forty items was as follows: six punctuation items, fifteen sentence structure items, eight grammar items, ten diction-and-

style items, and one logic-and-organization item. Each experimental unit and each passage-related unit was administered to approximately 300 examinees as part of the ACT Assessment tryout procedures.

The responses to each item were classified into a 2 X 2 X 3 contingency table based on whether the response was: correct or incorrect; from the passage or the discrete form; and from an answer sheet which was in the top 27%, middle 46% or bottom 27% according to total score. The three variables in the table will be referred to as Response (correct/incorrect), Form (passage/discrete), and Score Group (high/middle/low).

A separate loglinear analysis was performed on each item of each test (Tests A and B). Since there are twenty items on each test, a significance level of .05 ÷ 20 = .0025 was used to simultaneously control the significance level for a set of twenty items at approximately .05 (according to the Bonferroni inequality). Similarly, an overall .10 level was achieved by using .005 on individual items. The loglinear analysis allows for the study of interactions between pairs of variables in the contingency table, as well as an interaction among all three variables.

In this model, an interaction between Form and Score Group can be interpreted as an artifact of the sampling procedure. Such an interaction indicates that the proportions of the answer sheets that are from the passage and discrete forms vary over the three score groups.

An interaction between Response and Score Group is expected for each item since such an interaction indicates that the proportion of correct responses varies with the score group. This proportion is expected to increase from low to middle to high score group. It should be noted that the size of this

interaction can be interpreted as an indication of the discriminating power of the item.

The third two-variable interaction, that between Response and Form, is an indicator of an overall difference between forms on the proportions of correct and incorrect responses. Hence it should be concluded that the form of item (passage or discrete) does affect the difficulty level of the item for those items on which this interaction is significant.

A three-way interaction (between all three variables) exists when the between-form differences in proportions of correct responses varies over the three score groups. Thus the size of this interaction indicates the degree to which the item discriminability differs on the two forms.

Loglinear analysis typically involves fitting a series of models that are hierarchical in the sense that all models investigated can be expressed by placing restrictions on the parameters of models higher up in the hierarchy (Bishop, Fienberg & Holland, 1975). The highest model in the hierarchy, much like an anova model, contains a grand mean term, main effects of each variable, and all possible interactions. This model is referred to as the saturated model and it yields a likelihood-ratio chi squared statistic of zero because this model will always fit the data perfectly. For models lower in the hierarchy this statistic, G^2, provides a measure of goodness of fit, as do differences between the G^2 statistics for models at different levels of the hierarchy.

Figure 1 shows four models, each at a different level of the hierarchy, that were used in this study. Other models exist at the third and fourth levels, but they are not relevant. Also shown in the figure are symbols for

the likelihood ratio chi squared statistics for each model, and for differences between adjacent models in the hierarchy. In the figure, R represents the Response variable, S the score group variable, and F the Form Variable; pairs of letters (eg., RS) represent two variable interactions, while RSF represents the three-way interaction. Since loglinear models must always contain main effects of each variable present in an interaction, the three main effects are all present in models at the first three levels and their symbols are not shown. In the model at level 4, however, R must be shown to indicate that the main effect of the Response variable is present as well as the Score Group by Form (SF) interaction. The other two main effects are also present. The G^2 statistics shown between the levels represent differences. For example, $G^2(n/dif) = G^2(n) - G^2(dif)$ is the statistic for testing the difference in fit of the models at levels 3 and 4.

In this study, the following tests were examined for each item:

1. Differential discrimination: $G^2(dis) = G^2(dis/s)$ since $G^2(s) = 0$. This test provides a test that the three-way interaction is zero. As was mentioned above, this interaction is an indicator of differential discrimination.
2. Differential difficulty: $G^2(dif/dis)$ provides a test of whether the two forms differ in difficulty of the item.
3. Overall discrimination: $G^2(n/dif)$ is a test of the Response by Score Group interaction which, as was mentioned above, should exist if the item has discriminatory power.

Following Bonett and Bentler (1983), we define a null model containing

effects determined by the sampling design of the study. If $G^2(n)$ shows no lack of fit, then there is little point in investigating any of the tests mentioned above. Also, $G^2(n)$ is used to compute the overall discrimination statistic as shown above.

RESULTS

Test A

Table 1 shows the difficulty indices (proportions of correct answers) for each score group as well as for the combined group, and Table 2 shows the likelihood-ratio chi square (G^2) statistics for this test. Figure 2 is a joint plot of the score group proportions for the two forms. If there was no difference between the discrete and passage difficulties at each score group for an item, the plot for that item would consist of three points along the 45° line. An item whose plot is parallel to the 45° line has a constant difference in difficulty between the two forms across score groups (items easier on the passage form are above the 45° line and those easier on the discrete are below it). Items whose plots are not parallel to the 45° line have nonconstant discrete-passage difficulty differences across the three score groups. Hence these items differ in discrimination between the two forms. We must, of course, examine the significance tests along with the plots in making inferences. Items 3, 4, 12, 15, and 18 on Test A merit some discussion, according to the results of the loglinear analysis.

Item 3 has no significant (.10 simultaneous level) overall discriminating power and hence is a poor item. In the plot of Figure 2, it can be seen that it does discriminate somewhat on the passage form between the low and middle

score groups, but has a negative discrimination on that form between the middle and high score groups. Although this item does discriminate positively on the discrete form, the amount of discrimination between the middle and high score groups is rather low. Since the G^2 statistics for differential discrimination is nonsignificant (.10 simultaneous level), one must be cautious in inferring that it discriminates on one form and not on the other. The values of the difficulty levels used to make the plots are shown in Table 1.

Item 4 is easier overall on the discrete form than it is on the passage form. This can be seen in Table 1 as well as in the plot of Figure 2. Although item 16 approaches significance in the same direction as item 4, it is not significant by the criterion of .10 on the simultaneous test. Since the individual p-level is .0107 this item would be significant at the .02 level if tested alone rather than simultaneously with the other 19 items.

In the other direction, items 12, 15, and 18 are easier overall on the passage form than on the discrete form. Item 3 approaches significance ($p = .0078$) in this same direction but its poor discrimination (referred to above) between the middle and high score groups appears to have prevented it from reaching significance on the simultaneous test.

Test B

Tables 3 and 4 show the difficulty levels and likelihood ratio chi-squared statistics, respectively, for Test B. The difficulty levels for each score group on the two forms are plotted in Figure 3. Items 7, 8, 10, and 15 are the items on Test B that merit discussion.

Item 8 has a nonsignificant overall discrimination. The plot in Figure 3 shows that this item discriminates negatively between the low and moderate score groups on the discrete form of the test. Overall, its discrimination on the discrete form is very poor, but it does discriminate somewhat better on the passage form. As was the case with item 3 on Test A, however, the G^2 statistic for differential discrimination was nonsignificant, so a difference in discrimination between forms cannot legitimately be inferred (at the .10 simutaneous significance level).

According to the plot, item 7 discriminates negatively between the low and middle score groups on the passage form. Again, however, it must be pointed out that the differential discrimination G^2 is nonsignificant. Overall, item 7 was significantly easier (.05 simultaneous level) on the passage than on the discrete form.

Items 10 and 15 were significantly easier (.10 and .05 simultaneous levels, respectively) on the discrete form than on the passage form.

DISCUSSION

According to the results of the loglinear analysis, only seven of the forty experimental items approached significance. The content classification of these seven items was as follows: in Test A, there were two sentence structure items, one diction-and-style item, and one logic-and-organization item; in Test B, there were three sentence structure items.

Of these significant items, four were easier overall in the passage format than in the discrete-sentence format (items 12, 15, and 18 in Test A,

and item 7 in Test B). Three were easier overall in the discrete-sentence format than in the passage format (item 4 in Test A, and items 10 and 15 in Test B). In general, the items that were easier in the passage format needed a larger context (i.e., the passage) from which to judge the correct answer. For example, in the logic-and-organization item (item 15, Test A), only a larger context could make clear whether a certain concrete detail was necessary or irrelevant; and in the sentence structure item (item 18, Test A), consistency of person/viewpoint could be judged only from the larger context of the passage. Insofar as some content classifications often demand a larger context (for example, logic-and-organization and sentence structure), those classifications would be more susceptible to the impact of format.

In general, the items that were easier in the discrete-sentence format tended to be items that were totally independent of the passage--its reading load and rhetorical concerns. Instead, they tested a fairly explicit and narrow skill. For example, item 10 in Test B presents a clearcut comma splice that is apparent in the discrete sentence, independent of the passage. Likewise, item 15 in Test B presents a dangling participle, another language situation that can be presented adequately in a discrete sentence.

Whether a language situation can be presented adequately in a discrete sentence is an important concern. In at least two instances--item 5 in Test A and item 2 in Test B--when a seemingly apparent and unequivocal syntactic problem was presented in discrete-sentence format,a multiplicity of keys resulted. Item 5 required the placement of an adverbial modifier, but only the larger context of the passage could clarify the necessary position. And item 2 measured the punctuation of coordinate adjectives one of which, the

word "economic," had two meanings ("thrifty" and "related to economics") that only the passage could clarify sufficiently for the correct punctuation to be chosen.

Since none of the other thirty-three items, which were drawn from all content classifications (i.e., punctuation, grammar, sentence structure, diction-and-style, and logic-and-organization), approached significance according to the results of the loglinear analysis, this study seems to indicate that for most of those items, the format of presentation had little effect on the performance of examinees. The reading load of the passage seemed no more disadvantageous than at the sentence level. In discrete-sentence format, then, the efficiency of testing conceivably would be improved. However, and this is a significant qualification, items presented in discrete-sentence format tend to be successful only when they measure relatively narrow skills and only if the sentence context is sufficient to avoid ambiguity and, consequently, multiple keys. Items that measure more sophisticated rhetorical concerns will invariably demand the context of a passage. For those who are interested in measuring a range of skills in multiple-choice format and are interested in doing so in a manner perceived by language arts people to be more valid, the passage-related format would seem to be the format of choice.

REFERENCES

Bishop, Y.M.M., Fienberg, S.E., & Holland, P.W. (1975). Discrete multivariate analysis: theory and practice. Cambridge, MA: MIT Press.

Bonett, D.G., & Bentler, P.M. (1983). Goodness-of-fit procedures for the evaluation and selection of log-linear models. Psychological Bulletin, 93 149-166.

Godshalk, F.I., Swineford, F., and Coffman, W.E. (1966). The measurement of writing ability. NY: College Entrance Examination Board.

Huntley, R.M., Schmeiser, C.B., & Stiggins, R.J. (1979). The assessment of rhetorical proficiency: The role of objective tests and writing samples. Paper presented at the annual meeting of the National Council on Measurement in Education, San Francisco, CA.

Huntley, R.M. (1977). What's REALLY basic in the language arts? Iowa City, IA: American College Testing Program.

Table 1. Difficulty Indices: Test A

Item	Low 27%		Mid 46%		Hi 27%		Overall	
Number	Discr.	Pass.	Discr.	Pass.	Discr.	Pass.	Discr.	Pass.
1	.436	.351	.821	.743	.984	.934	.739	.709
2	.359	.238	.406	.478	.574	.569	.433	.440
3	.154	.317	.255	.373	.295	.333	.233	.348
4	.308	.197	.453	.318	.525	.369	.424	.299
5	.397	.429	.632	.539	.738	.646	.584	.539
6	.692	.696	.858	.826	.967	.969	.833	.826
7	.667	.730	.868	.853	1.000	.934	.837	.846
8	.833	.694	.868	.909	.984	1.000	.886	.879
9	.808	.852	.962	.936	.984	.972	.918	.926
10	.333	.458	.632	.688	.885	.810	.600	.659
11	.269	.366	.585	.579	.754	.759	.526	.559
12	.141	.413	.387	.496	.689	.692	.384	.527
13	.397	.423	.736	.698	.902	.819	.669	.653
14	.308	.220	.604	.616	.820	.793	.563	.559
15	.462	.645	.651	.860	.770	.891	.620	[redacted]
16	.128	.085	.264	.170	.410	.259	.257	.170
17	.667	.698	.887	.863	1.000	.947	.845	.846
18	.295	.500	.349	.653	.410	.719	.347	.632
19	.115	.062	.132	.127	.311	.373	.171	.178
20	.679	.721	.915	.936	.967	.944	.853	.884

Table 2. Likelihood-ratio Chi-squared Statistics: Test A

Item No.	Overall Discrimination G^2(n/dif) df = 2	p-lev	Differential Difficulty G^2(dif/dis) df = 1	p-lev	Differential Discrimination G^2(dis) df = 2	p-lev
1	116.25	<.0001**	4.03	.0448	1.02	.5997
2	19.55	.0001**	.01	.9473	3.61	.1643
3	4.00	.1354	7.07	.0078	1.78	.4114
4	11.33	.0035*	9.54	.0020**	.02	.9912
5	22.04	<.0001**	1.69	.1949	1.66	.4350
6	40.47	<.0001**	.16	.6897	.28	.8693
7	40.45	<.0001**	.04	.8367	6.77	.0340
8	37.28	<.0001**	.33	.5635	5.90	.0524
9	23.39	<.0001**	.01	.9262	1.46	.4813
10	62.04	<.0001**	1.03	.3106	3.21	.2011
11	55.87	<.0001**	.45	.5027	1.18	.5544
12	50.77	<.0001**	9.00	.0027*	7.10	.0287
13	68.67	<.0001**	.58	.4435	1.74	.4195
14	81.75	<.0001**	.39	.5345	1.10	.5767
15	30.49	<.0001**	20.88	<.0001**	.85	.6529
16	19.80	.0001**	6.50	.0107	.12	.9436
17	48.60	<.0001**	.26	.6099	4.99	.0826
18	9.61	.0082*	38.78	<.0001**	.91	.6346
19	32.50	<.0001**	<.01	.9623	1.83	.4003
20	45.80	<.0001*	.26	.6098	.77	.6818

**p < .0025: Simultaneous tests of 20 items significant at .05 level

*p < .005: Simultaneous tests of 20 items significant at .10 level

Table 3. Difficulty Indices: Test B

Item Number	Low 27%		Mid 46%		Hi 27%		Overall	
	Discr.	Pass.	Discr.	Pass.	Discr.	Pass.	Discr.	Pass.
1	.422	.318	.391	.391	.577	.600	.455	.427
2	.494	.460	.609	.557	.789	.662	.622	.560
3	.759	.683	.946	.833	.958	.974	.886	.838
4	.482	.467	.641	.689	.873	.815	.654	.667
5	.386	.397	.543	.557	.718	.723	.541	.560
6	.627	.541	.641	.743	.845	.847	.695	.723
7	.253	.525	.380	.446	.465	.724	.362	.537
8	.386	.328	.272	.376	.338	.569	.329	.421
9	.699	.667	.848	.739	.859	.892	.801	.761
10	.157	.129	.370	.215	.676	.531	.386	.275
11	.554	.521	.750	.701	.859	.845	.715	.682
12	.361	.333	.500	.500	.620	.605	.488	.490
13	.422	.500	.598	.653	.746	.813	.581	.656
14	.361	.354	.413	.536	.606	.493	.451	.475
15	.386	.169	.543	.393	.746	.517	.549	.367
16	.241	.423	.457	.469	.803	.708	.484	.527
17	.639	.730	.902	.861	.972	.969	.833	.856
18	.361	.231	.543	.418	.732	.701	.537	.446
19	.313	.305	.380	.429	.479	.552	.386	.428
20	.735	.683	.870	.809	.944	.810	.846	.777

Table 4. Likelihood-ratio Chi-squared Statistics: Test B

Item No.	Overall Discrimination G^2(n/dif) df = 2	p-lev	Differential Difficulty G^2(dif/dis) df = 1	p-lev	Differential Discrimination G^2(dis) df = 2	p-lev
1	16.52	.0003**	.33	.5684	1.44	.4877
2	18.58	.0001**	2.39	.1222	1.13	.5696
3	37.85	<.0001**	4.38	.0363	3.37	.1852
4	44.31	<.0001**	<.01	.9741	1.40	.4969
5	31.69	<.0001**	.05	.8159	<.01	.9980
6	24.12	<.0001**	.20	.6588	3.31	.1911
7	13.94	.0009**	15.84	.0001**	5.10	.0781
8	5.79	.0553	4.70	.0302	6.09	.0475
9	15.24	.0005**	1.56	.2116	2.65	.2662
10	71.64	<.0001**	8.48	.0036*	.85	.6525
11	34.53	<.0001**	.74	.3909	.08	.9598
12	20.69	<.0001**	.08	.7750	.08	.9630
13	31.63	<.0001**	2.31	.1281	.09	.9537
14	11.16	.0038*	.14	.7121	4.72	.0944
15	33.53	<.0001**	18.71	<.0001**	1.33	.5130
16	59.42	<.0001**	.64	.4223	6.89	.0319
17	50.39	<.0001**	.08	.7745	2.15	.3413
18	50.69	<.0001**	5.41	.0200	.90	.6369
19	11.47	.0032*	.73	.3938	.45	.7991
20	13.61	.0011**	4.97	.0258	2.72	.2562

**$p < .0025$: Simultaneous tests of 20 items significant at .05 level

*$p < .0050$: Simultaneous tests of 20 items significant at .10 level

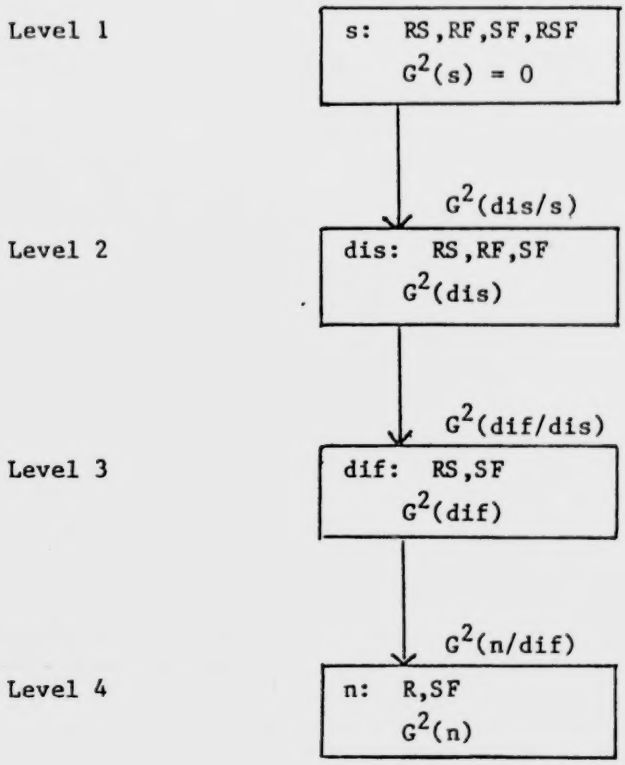

Figure 1. Models in the Hierarchy

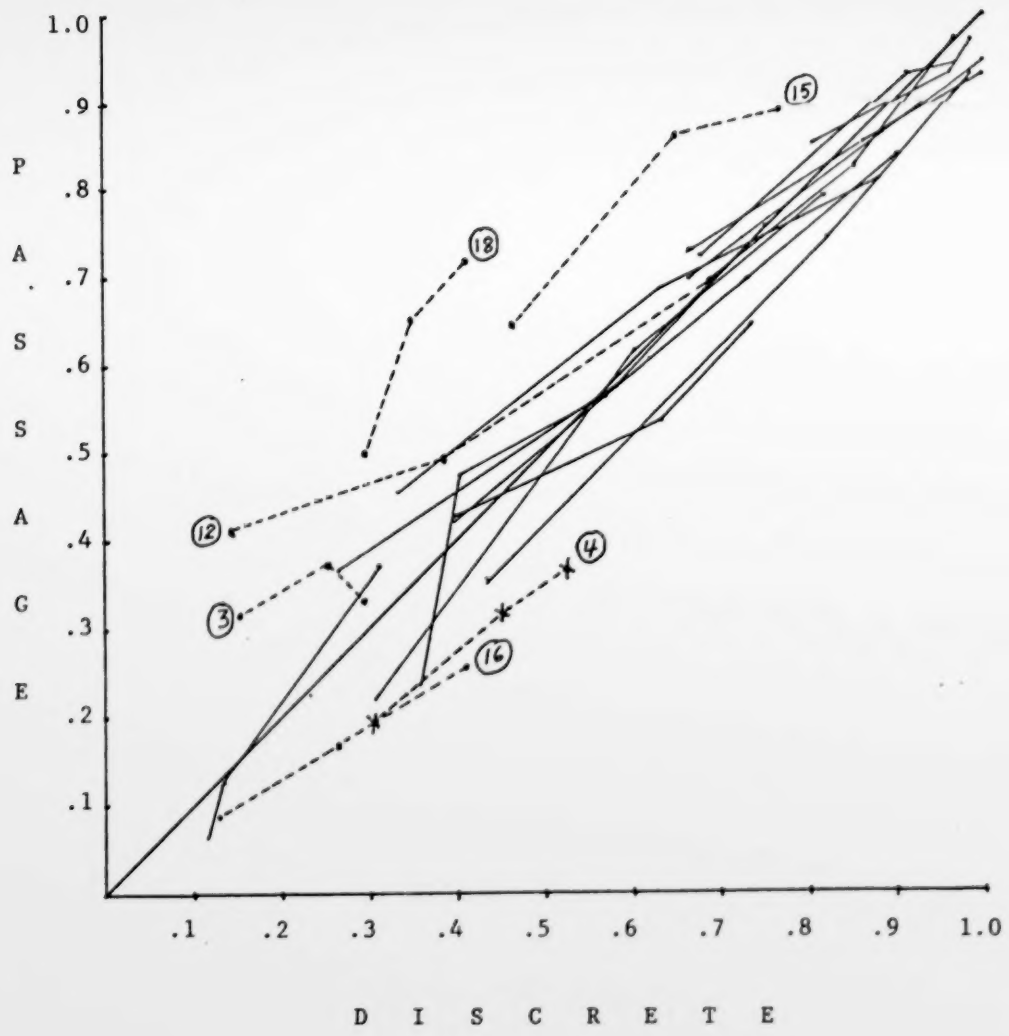

Figure 2. Test A Difficulty Levels by Score Groups

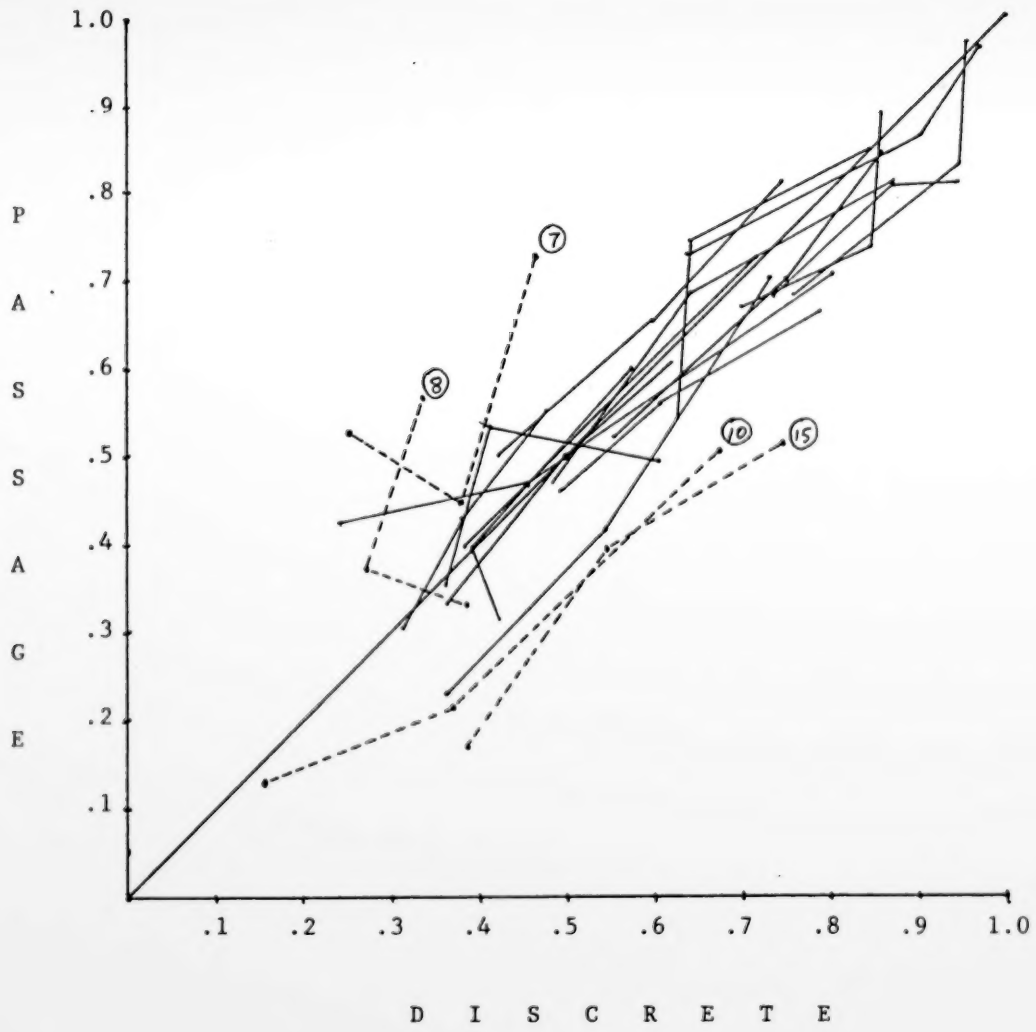

Figure 3. Test B Difficulty Levels by Score Group

Passage Presentation

Most people only think of salt as a seasoning
1

for food. It is, of course, a primary flavoring
2

throughout the world in different dishes,
3
but less than five percent of all salt

produced finds its way into saltshakers.
4
Indeed, it has been estimated that there

are over 14,000 or more uses for this mineral.
5

Salt and its by-products, are used by
6

meat packers, chemical companies, hide and

leather processors, and in food processing.
7

Where does all this salt come out of? Well,
8

1. A. NO CHANGE
   B. (Place after *think*)
   C. (Place after *of*)
   D. (Place after *salt*)
   E. (Place after *for*)

2. F. NO CHANGE
   G. moreover,
   H. nevertheless
   J. however,
   K. on the contrary,

3. A. NO CHANGE
   B. in dishes throughout the world that are different,
   C. in different kinds of dishes in the world,
   D. throughout the world,
   E. throughout the world in dishes of different kinds,

4. F. NO CHANGE
   G. find its
   H. find it's
   J. finds it's
   K. find the

5. A. NO CHANGE
   B. at least
   C. approximately
   D. nearly
   E. OMIT the underlined portion.

6. F. NO CHANGE
   G. Salt, and its by-products
   H. Salt, as well as its by-products,
   J. Salt along with its by-products
   K. Salt and its by-products

7. A. NO CHANGE
   B. food processors.
   C. by food processors.
   D. for food.
   E. food processing.

8. F. NO CHANGE
   G. come?
   H. come out?
   J. come out from?
   K. come from?

Discrete-Sentence Presentation

1. There are, in general, three ways that people react to something new and unfamiliar: with ridicule, or suspicion, or with curiosity.

   A. NO CHANGE
   B. suspiciously,
   C. showing suspicion,
   D. being suspicious,
   E. with suspicion,

2. The Arabs say, "There is salt between us," meaning, "We have eaten together and are friends."

   F. NO CHANGE
   G. It is said by Arabs,
   H. A saying by Arabs is,
   J. In Arabian they say,
   K. In Arab countries, Arabs will say.

3. Unlike dark caves and rabbit holes, the atmosphere of underground houses can be bright and airy.

   A. NO CHANGE
   B. caves' and rabbit holes', the atmosphere of
   C. caves and rabbit holes, underground houses
   D. caves and rabbit holes the atmosphere of underground houses
   E. caves and rabbit holes', the underground house's atmospheres

4. It is true that in England, as Eliza Doolittle declares in *My Fair Lady*, "hurricanes hardly happen."

   F. NO CHANGE
   G. In England,
   H. All over England,
   J. It's a fact that in England,
   K. From one end of England to another,

5. Most people only think of salt as a seasoning for food.

   A. NO CHANGE
   B. (Place after *think*)
   C. (Place after *of*)
   D. (Place after *salt*)
   E. (Place after *for*)

6. When I was a boy, baseball fever appears simultaneously with the first warm days of spring.

   F. NO CHANGE
   G. was appearing
   H. appeared
   J. had appeared
   K. has appeared

7. Their bills are so powerful that some species, such as the macaws, have easily cracked the hardest walnuts.

   A. NO CHANGE
   B. can crack with ease
   C. can easily crack
   D. are easily cracking
   E. crack

8. Most of us feel profound pleasure in simply getting around the ocean.

   F. NO CHANGE
   G. hanging around
   H. being near
   J. getting next to
   K. being in close proximity to

9. On your way back down, if you want to proceed more quickly, just slide down the snowfield in what is known as a "glissade."

   A. NO CHANGE
   B. quicker,
   C. quick,
   D. swift,
   E. speedy,

10. Making a Saturday trip to the dump, used to be a tradition in small Maine towns.

    F. NO CHANGE
    G. dump used to be
    H. dump, often used to be
    J. dump would have been
    K. dump; it used to be